



LSTM 실험 Protocol

Tae-Won Lee,
leetee95@gmail.com
2020. 02. 02.

Experiment

- Dataset

stocks	N	Train	val	test
KOSPI 200 INDEX				
Gold price				
Oil price				
CB interest rate(AA-Grade 회사채)				
KTB interest rate(한국 국고채)				
KOSPI 200 INDEX log difference				

- (2000.01.01 – 2020.12.31) 기간 내의 데이터.

- 각 데이터는 시초가(Open), 고가(High), 저가(Low), 종가(Close), 수정주가(Adj Close), 거래량(Volume) 6개의 변수가 존재.

- > 차후 실험에서 종가(close)를 사용하여 예측.

Experiment

- Statistics

Table 1

Summary statistics for explanatory variables: *** and ** denote a rejection of the null hypothesis at the 1% and 5% significance level, respectively.

	Mean	Standard deviation	Skewness	Kurtosis	Jarque-Bera	ADF
KOSPI 200 INDEX	189.47	68.18	-0.47	1.73	416.02***	-1.63
Gold price	906.65	463.03	0.14	1.72	282.50***	-1.30
Oil price	64.65	27.79	0.21	2.02	186.29***	-1.85
CB interest rate	4.56	1.55	-0.02	2.42	55.45***	-1.61
KTB interest rate	3.92	1.32	-0.20	2.27	115.74***	-1.74
KOSPI 200 INDEX log difference	0.00015	0.0065	-0.39	8.81	5675.93***	-62.23***
ARCH-LM Test (Lag = 5)			F-statistic	2.236**		
			Obs*R-squared	11.17**		

- ARCH-LM Test 제외

-> ARCH, GRACH 등의 모형에 적용 가능 여부 판단. RNN, LSTM모형에서는 제외.

Experiment

- Experiment Settings

Hyperparameters	Number
hidden state, cell state dimension	10
input sequence length	7
output sequence length	1
epoch	15
Batch size	128

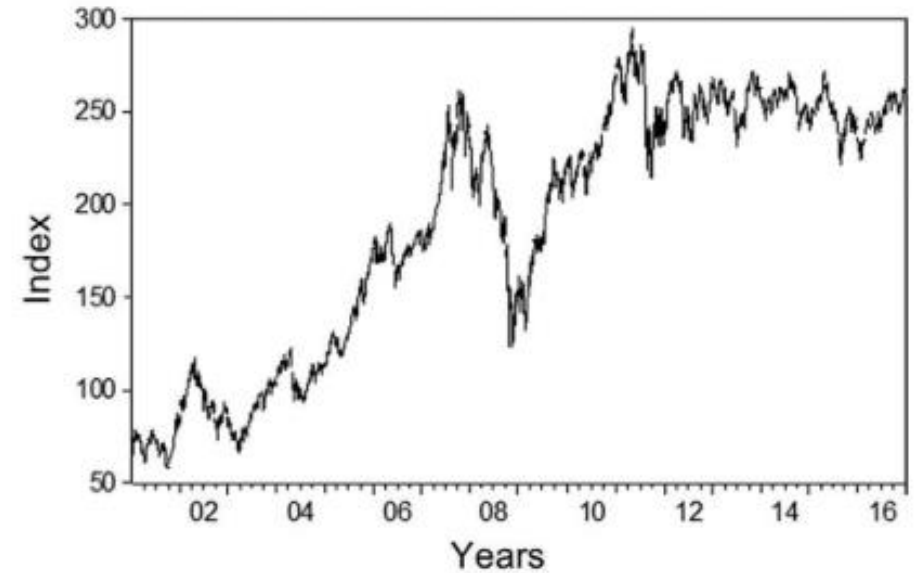


Fig. 1. KOSPI 200 index.

- 7 시점 간의 과거 자료로 그 다음 한 시점의 주가를 예측하는 모형 생성

Experiment

- logarithmic return

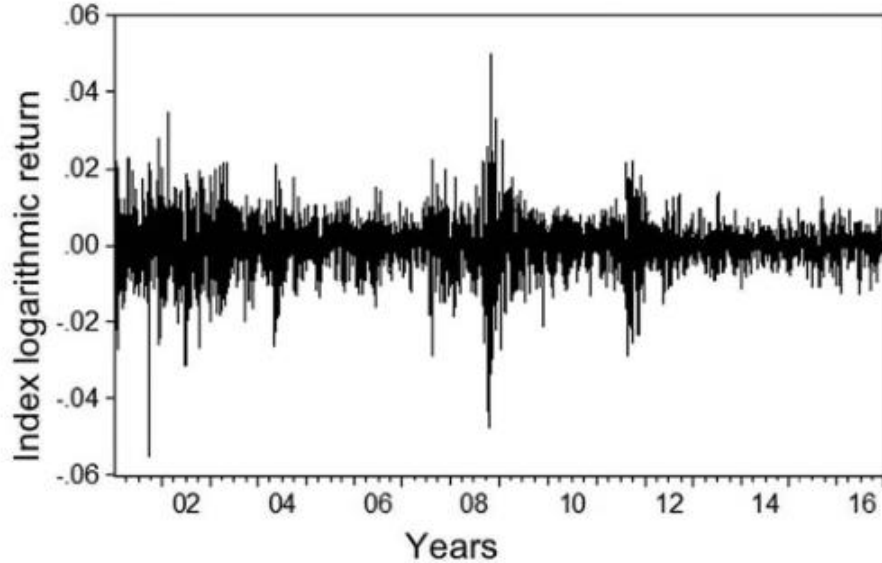


Fig. 2. KOSPI 200 index logarithmic return.

- Histogram

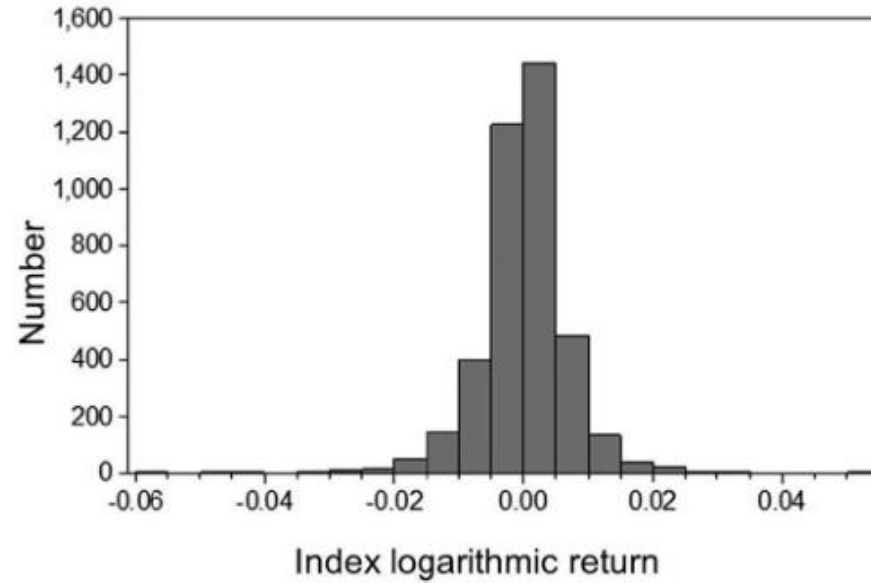


Fig. 3. Log difference in KOSPI 200 index histogram.

- logarithmic return : $\ln((t+1)/t)$ 로 나타냄

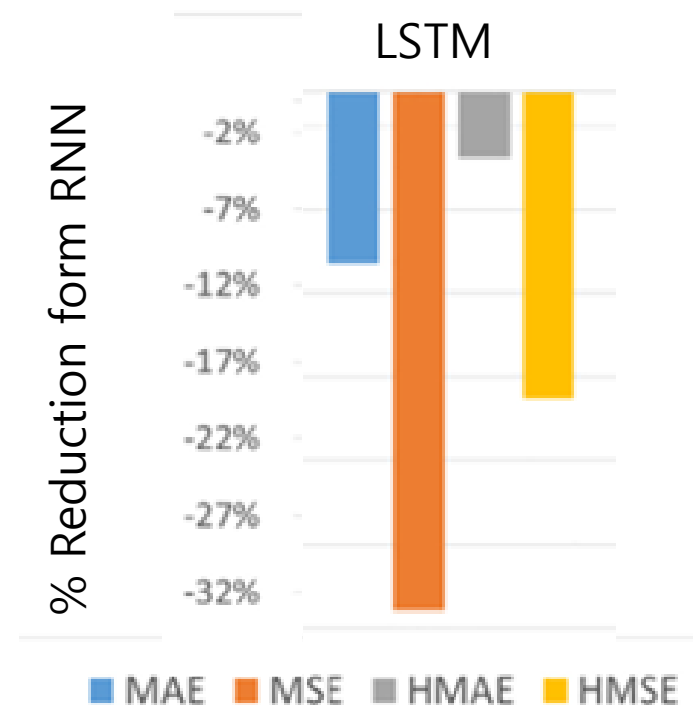
t : 시점이 t일때의 종가(close)

Experiment

- Results of out-of-sample forecasts with various loss function

Model	MAE	MSE	HMAE	HMSE
RNN				
LSTM				

- 최종 모형에서 Test 자료의 loss function 계산



Experiment

- Realized volatility forecast results

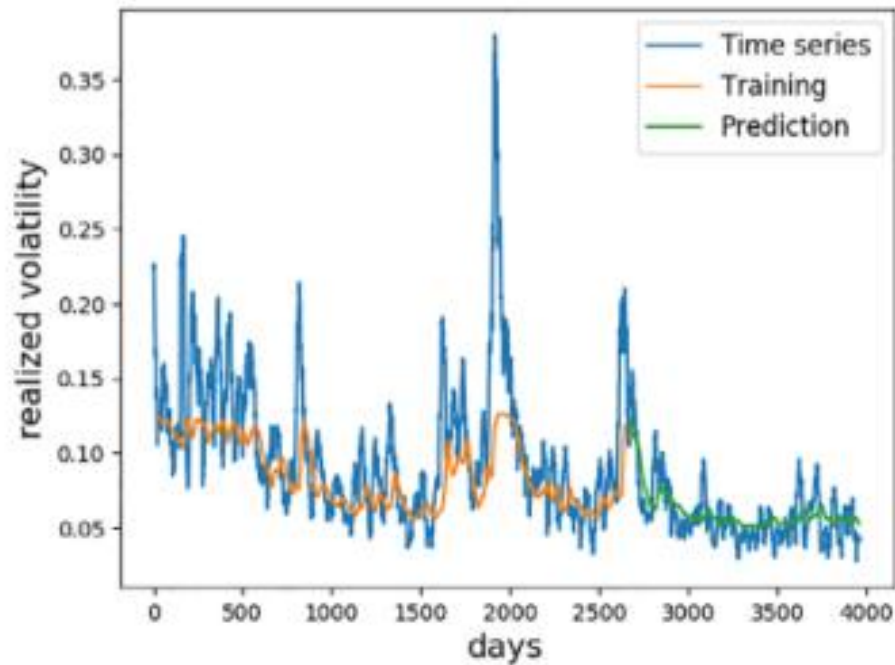


Table 6

Comparison of out-of-sample forecasts errors over different time horizons for various days-ahead prediction of volatility: Errors are calculated by HMAE

	1 day ahead prediction Windows length (days)			14 days ahead prediction Windows length (days)			21 days ahead prediction Windows length (days)		
	7	15	22	7	15	22	7	15	22
DFN	0.6825	0.6764	0.6598	0.7056	0.6907	0.6753	0.7177	0.7009	0.6770
LSTM	0.5559	0.5289	0.5186	0.5617	0.5331	0.5270	0.5889	0.5514	0.5474
W-DFN	0.6651	0.6410	0.6363	0.6860	0.6673	0.6504	0.6953	0.6892	0.6629
G-DFN	0.6326	0.6130	0.6086	0.6541	0.6362	0.6204	0.6837	0.6672	0.6450
E-DFN	0.6111	0.5878	0.5700	0.6346	0.6122	0.6080	0.6524	0.6323	0.6282
W-LSTM	0.5096	0.5040	0.4912	0.5572	0.5139	0.5100	0.5722	0.5508	0.5108
G-LSTM	0.5046	0.5001	0.4526	0.5511	0.5095	0.4752	0.5629	0.5393	0.5055
E-LSTM	0.5014	0.4633	0.4465	0.5355	0.4819	0.4523	0.5363	0.4861	0.4715
GW-LSTM	0.4990	0.4459	0.4419	0.5249	0.4485	0.4458	0.5331	0.4485	0.4472
EW-LSTM	0.4653	0.4375	0.4340	0.4851	0.4398	0.4381	0.4919	0.4399	0.4398
GE-LSTM	0.4498	0.4331	0.4328	0.4637	0.4375	0.4357	0.4774	0.4395	0.4389
GEW-LSTM	0.4359	0.4305	0.4291	0.4509	0.4364	0.4343	0.4597	0.4376	0.4349

- Realized volatility : 과거의 데이터를 이용해 변동성을 계산한 값.

Reference

- 1 Ha Young Kim, Chang Hyun Won (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. (2018)

Appendix

Appendix

- Jarque-bera test
 - 왜도와 첨도가 정규분포와 일치하는지 판단하는 적합도 검정
- ADF test
 - 시계열 데이터가 정상성(stationary)을 만족하는지에 대한 검정
- LM-ARCH 검정
 - ARCH, GARCH와 같은 모형을 적용할 수 있는지에 대한 검정.